MULTIMEDIA UNIVERSITY®

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2022/2023 (TERM ID 2215)

HPB2021 –BIOINFORMATICS PROGAMMING II

(All sections / Groups)

13 FEBRUARY 2023
9.00 am – 11.00 am
(2 hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 3 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal marks (10).
3. Write all your answers in the Answer Booklet provided.
4. You may use a calculator.

Question 1

a) Write a Python code to [6 marks]
   i) accept a file as a user input
   ii) reverse complement a FASTA sequence
   iii) translate the reverse complemented sequence using dictionary

b) Read all the sequences in a FASTA formatted file "example.fa" using Python. Display the longest sequence, ID of the sequence and the length [4 marks]

Question 2

a) The following is the Python code on a bank account. Fill in the code marked with #A to #F [10 marks]

```
class BankAccount:
  bank_name = "CBM BHD"
  def __init__(self, name, ID, bal=0):

    #A

  def withdraw(self, amount):

    #B – update balance

  def deposit(self, amount):
    #C – update balance

 def __str__(self):

    #D

class MinimumBalanceAccount(BankAccount):
  def __init__(self,name, ID, bal, minimum_balance):

    #E

  def withdraw(self, amount):

    #F

x = MinimumBalanceAccount("sofi", 123, 500)
print (x)
print (x.withdraw(300))
```

Continued .....

Question 3

a) Write BioPython code to create a Seq object of DNA using the IUPAC.unambiguous_dna alphabet to find the percentage of 'G' and 'C' nucleotides in the sequence. Then calculate the percentage of 'A' and 'T' in the sequence.

[5 marks]

b) Read all the sequences in a fasta file "example.fasta" using BioPython. Write the sequences equal or longer than 200bp into a new file, namely "example_200.fasta".

[5 marks]

Question 4

a) Given x<-matrix(c(3,4,-2,6),nrow=2,byrow=T) and y<-matrix(c(1,3,0,9,5,-1),nrow=3,byrow=T), write the output for [5 marks]

- x
- y
- x*y
- y*x
- y%*%x

b) Write an R program to accept DNA sequence from user. The program counts the number of bases, "A","T","C", and "G", prints the length of the sequence, bases, and the number of occurrences. [5 marks]

END OF PAPER